Table I about here

Similar relative frequency hierarchies have been observed by others and are probably consistent with the experience of most teachers at the primary and intermediate levels. Ackerman (1972) describes a phenomenon which exactly parallels the Funtion Rater hierarchy, referring to it as "a short chain which has occurred in every classroom (p. 41)."

In sum, two hierarchies have been identified, Relative Frequency and Precedence. The precedence hierarchy reverses relative frequency, assigning the highest precedence to the lowest frequency, next highest to next lowest, and so on.

## OPERATIONALIZING FUNCTIONAL CATEGORIES

The effectiveness of the system depends upon the adequacy of the operational definitions by which observers recognize the various categories. One advantage of working with relatively common descriptors is that they are widely understood even before training of observers is begun. Forty-seven undergraduate special education students, naive with respect to the Function Rater, were given the words relevant, unproductive, disruptive, aggressive, and asked to write the critical characteristics by which they could recognize classroom behavior possessing these attributes. Overall, 80% of the student definitions were judged to be in essential conformity with these basic definitions:

Relevant - What the teacher says to do

Unproductive - Not what the teacher says, but not bothering anyone else

Disruptive - Interrupting the work of others

Aggressive - Attacking the person or property of others

A large proportion of classroom behavior can be reliably rated with little more information than the brief definitions shown above. Further refinement is possible with adoption of the additional operational guidelines which are presented in detail in Appendix A (The Delaware Function Rater: Guiding Concepts for Ratings). For present purposes, it is enough to say that the fundamental criterion for classifying discrete behaviors into functional classes is the Rule of Probable Effect:

> The probable effect of an interval of behavior is the effect it would produce if its least adaptive component occured continually or repeatedly.

Classification based on probable effect avoids the difficulties of categorizing behavior on the basis of intention, which cannot be reliably inferred, or actual effect, which may not reflect, in a single instance, the way an operant generally functions. A child who throws an object at a classmate may not produce the usual effect if he misses the target and the teacher does not see him. However, many repetitions of the behavior would probably lead to his being treated by the social environment as an aggressive child—one to be avoided, guarded against, and, ultimately, isolated. In judging the probable effect of a given behavior, Raters ask the question, "How would the objectives of education be served for this child if he behaved this way all of the time?"

## TECHNICAL CHARACTERISTICS

### Reliability

The reliability of the method has not been thoroughly evaluated, although preliminary studies suggest that high levels of inter-rater reliability can be obtained if, at the beginning of a rating project, raters are allowed to collaborate on conventions to cover the idiosyncrasies of a particular

classroom setting. A relatively controlled attempt to assess reliability was made for the fourth set of data shown in Table I (1691 cell-by-cell agreements, 166 disagreements, percent agreement = 91.1). These ratings were taken during training of students by the authors. Agreements between individual raters within this group ranged from 82% to 100%. Other reliability figures given in Table I were obtained by students working without benefit of direct supervision.

Inter-rater reliability is a matter of concern if the ratings are used as experimental dependent measures, but of less importance if the Function Rater is used, as it has been so far, for purposes of teacher training. One caution should be observed by anyone contemplating the use of this or similar systems for purposes of making comparisons across environments. Discriminations of categories are based in part on the rule structure of a given environment. Reliabilities are a function of the adequacy with which the rule structure yields clear operational demarcations of specific behaviors. The problem with comparisons between environments is that a single set of operational definitions may not be equally reliable in more than one setting. On the other hand, different sets of definitions would confound the comparison. Rating systems of this kind are most appropriately used in a within-setting paradigm and not between settings.

## Validity

Traditional concern for concurrent, construct and content validity is inappropriate in a system that measures classroom behavior directly. Coding systems are descriptive; hence, their validity does not depend on their relatedness to external criteria. One might argue for a change in definitions or the inclusion of a particular behavior in a category other than the one

suggested in rating conventions, but once the definitions are accepted, the only remaining question is whether they have any significant meaning (i.e., whether they possess predictive validity). For example, it may be questioned whether the level of a child's relevant behavior ("what the teacher says to do") is related in an important way to his achievement; in other words, whether "relevant" is relevant. The Function Rater makes no prediction on the matter but treats it as a purely empirical question. The best that a behavior coding system can do is make the answer more accessible by providing a reliable measurement of the relative frequency with which Task Relevant behavior occurs.

Sensitivity

Instrument sensitivity is a dimension of special importance in a system that reduces all behavior to only four mutually exclusive categories. If a system of the highest sensitivity were to be developed, it would probably provide for coding of behaviors according to both function and topography (e.g., relevant out-of-seat versus unproductive or disruptive out-of-seat). The observer would operate a multi-channel event recorder, depressing and releasing keys for the onset and offset of behaviors, thus providing a record of durations as well as frequencies. Containing a precise record of all available information, such a system would be sensitive to even the slightest behavioral changes.

By contrast, the Function Rater sacrifices much of this information--first, by ignoring response topography and classifying only on the basis of function (probable effect); second, by ignoring duration within intervals; third, by allowing only one descriptor per interval. The question is whether the information that survives the filtering process is capable of reflecting

behavioral changes that are related to changing environmental conditions. The affirmative evidence on this point is outside the scope of this report and will be presented in a separate article.

## THE MECHANICS OF RATING

A completed Function Rater form is shown in Figure 1. Rows represent successive minutes in a 20 minute observation period. The block of columns to the left are divided into five segments, each representing a 10-second interval. The last 10-second interval of each minute is not rated but used by the rater for changing positions, writing comments, etc. The four columns to the right are not used until the conclusion of the rating period. Then the number of R's, U's, D's and A's in each row are counted and entered in the appropriate columns to the right; e.g., during the first minute, there were 2 relevant segments, 2 unproductive, and 1 disruptive. When all the ratings have been counted and recorded in this manner, the numbers in the columns on the right are totalled and entered in the boxes lower right. If all twenty minutes have been rated, the total number of segments will be 100, and the figures in the boxes lower right will reflect relative frequency in percentages. If less than the full 20 minutes have been rated, percentages are calculated by dividing the total in each category by the total for all categories. These figures are then entered along the bottom line of the boxes lower right.

### Recording Plans

Behavior ratings may be taken on individuals or groups. The two basic options may be expressed as (1) twenty successive minutes rated on one child or (2) twenty children rated successively for one minute each. Any number of

combinations within these extremes are possible; e.g., a group of five children could be rated for four minutes each (four times at one minute). The principal restriction on a group rating of any kind is that the sequence of observations be planned in advance. For example, if the chairs are arranged in rows, the first day's ratings might start with a one minute sample of the child in chair 2, row 3; thence, chairs 3, 4, 5 and 6 in the same row, alternating up and down rows. Schemes followed in subsequent days would be different again. The importance of following a random sequence cannot be overstated. Failure to do so can result in selection biases of the kind that occur when, for example, observers choose to rate the most exotic forms of behavior going on at a given moment.

In general, an effort should be made to perform ratings at about the same time every day, preferably during a period that is devoted to the same kind of activity from day to day. Independent seatwork is probably the easiest milieu to rate, particularly at the lower grades.

## DISCUSSION

Qualitative descriptions of classroom behavior are so firmly entrenched in the language of special education that the possibility of improving the language in fundamental ways is seldom considered. There is nothing wrong with describing children as lethargic, for example; as distractible, hyperactive or emotionally labile. Terms such as these evoke highly specific images of the kinds of behavior they describe. It is when comparisons or assessments of change are attempted that the limitations of qualitative language become evident--for example, comparisons of a child's distractibility under various reinforcement conditions; or the change in emotional

lability from September to March. The weakness of qualitative language is exposed in such commonplace expressions as "He's showing improvement," or "She's not as bad as she was when she first came here." One wants to know how much improvement--how bad the child was, and how bad she is now.

What does the teacher mean when she says that a child is "constantly in motion" or "always yelling out?" Surely constantly and always are exaggerations, but the question remains: Does always mean 90% of the time, 50%, or only 12%. If it is 12% but no other pupil in the class comes anywhere near that relative frequency, it may be perceived as always. Of course it is not, and it is important to know that it is not.

It is even more important that teachers acquire the habit of reducing complex behavioral phenomena to manageable proportions. Behavioral problems that are viewed only in qualitative terms are much more difficult to treat than problems that have been measured. The teacher who knows only that a child is making a shambles of the class knows much less than the one who knows that the child is disruptive 38% of the time during programmed reading. The goal of the first teacher is survival; the goal of the second teacher is to reduce the relative frequency of disruptive behavior to 32% by Christmas and 24% by the end of the school year.

The argument for quantitative language holds for any discipline concerned with the management of behavior, but particularly so for special education. The teacher of exceptional children must often work with small increments of improvement over substantial periods of time. If not measured, they are likely to go unnoticed. As Ackerman has observed, "Behavior changes so slowly and steadily that it is like the growth of children: You don't

notice it unless you are away for a while, or try on last year's coat or shoes (1972, p. 12)."

From the standpoint of teacher training, the task does not end with presenting the case for quantitative language. The real problem is to change teacher behavior in ways that endure after the teacher leaves the training situation. In many cases, this turns out to be surprisingly difficult. As beguiling as the arguments for direct behavioral measurement may be, the effect of these arguments on day-to-day teaching practice would have to be judged negligible at present. One can speculate that teachers have lifetime histories of being reinforced for thinking of behavior exclusively in qualitative terms. If this is the case, the transition to a measurement oriented nomenclature may continue to be slow in coming.

The Function Rater is designed to speed the process along. It is neither the most "scientific" of rating systems nor the simplest.[1] But it does speak to teachers in a language they understand about things they wish to know.

[1]Kubany and Sloggett (1973) have suggested a variable interval coding system that yields data similar to that of the Function Rater but can be collected by the teacher without the help of external observers. It differs from the Function Rater in that it does not provide for continuous observation, is slightly less sensitive to the more maladaptive behavior classes, and requires longer sampling periods, thus limiting the specificity of the data (longer sampling periods would make it difficult to relate the data to a single activity such as a 30 minute arithmetic seatwork assignment). However, the advantages of the system outweigh the limitations. A student trained in systematic observation with the Function Rater would be able, as a teacher, to carry the basic idea forward in the classroom with a minimum of interference with other duties.

## REFERENCES


Ackerman, J. Operant Conditioning for the Classroom Teacher, Glenview, Ill.: Scott, Foresman and Co., 1972.

Becker, W., Madsen, C., Arnold, M. and Thomas, D. The contingent use of teacher attention and praise in reducing classroom behavior problems, Journal of Special Education, 1967, 1(3), 287-307.

Ferster, C. and Perrott, M. Behavior Principles, New York: Meredith Corp., 1968.

Kubany, E. and Sloggett, B. Coding procedure for teachers, Journal of Applied Behavior Analysis, 1973, 6(2), 339-344.

Madsen, C., Becker, W., and Thomas, D. Rules, praise, and ignoring: Elements of elementary classroom controls, Journal of Applied Behavior Analysis, 1968, 1(2), 139-150.

O'Leary, K. and Becker, W. Behavior modification of an adjustment class: A token reinforcement program, Exceptional Children, 1967, 33, 637-6[illegible].

O'Leary, K., Becker, W., Evans, M., and Saudargas, R., A token reinforcement program in a public school: A replication and systematic analysis, Journal of Applied Behavior Analysis, 1969, 2(1), 3-13.

Simon, A. and Boyer, E. (Eds.). Mirrors for Behavior: An Anthology of Observation Instruments, Philadelphia: Research for Better Schools, Inc., 1967.

Thomas, D., Nielsen, L., Kuypers, D. and Becker, W. Social reinforcement and remedial instruction in the elimination of a classroom behavior problem, Journal of Special Education, 1968, 2(3), 291-305.

APPENDIX A

THE DELAWARE FUNCTION RATER:
GUIDING CONCEPTS FOR RATINGS

It is helpful to discriminate two broad classes of behavior: (1) active, publicly-observable behaviors; and (2) passive states, in which little movement is discerned. Fortunately, active behaviors--the easier of the two to judge--comprise the majority in the repertoires of most children.

Active Behaviors

Activities expected of children engaged in academic tasks include reading aloud, speaking, writing, coloring, drawing, pasting, using instructional devices, and a host of others. Two basic criteria for the relevant rating are:

(1) that the behavior conforms to the teacher-directed task; e.g., speaking is relevant if the child has been asked to recite but may be disruptive under other circumstances.

(2) that the behavior is related to material the teacher has assigned; e.g., completing dot-to-dot puzzles is relevant if assigned but otherwise unproductive.

Passive States

For all practical purposes, passives will be rated either as Relevant or Unproductive. Rarely will a passive be rated as Disruptive or Aggressive--the single exception, perhaps, being the rare event in which a child passively resists a direct teacher command. Examples of behaviors which involve little discernible movement but which may be nonetheless relevant, include silent

reading, listening, watching and thinking. Staring at a book, thinking about matters unrelated to the lesson, and daydreaming--all unproductive--may be topographically indistinguishable from relevant passives. When this is the case the child is given the benefit of the doubt and the interval is rated Relevant. Usually, the observer will be aided by the presence of two additional kinds of information in making this most difficult of rating judgments:

(1) Specific Clues - Facial expressions, eye movements and hand gestures often suggest the content of a child's thoughts. If a child looks away from his arithmetic assignment, then touches thumb to fingers as in counting, continued engagement in relevant activity is suggested. If his eyes wander from object to object about the room, non-attendance to the task may be more strongly suspected. Eyes that do not move in the pattern characteristic of silent reading are unproductive eyes if silent reading is the task. A facial expression that cannot be associated with what the teacher has said suggests that its owner has attended to something other than her presentation.

(2) Posture and Orientation Variables - It is true that relevant academic behavior can take place in a learner who is neither seated erectly nor oriented toward the front of the room. It is probably also true that the instructional environment interprets flagrant departures from standard posture and orientation as counter-productive. How the environment perceives these variables is the key to rating them. In general, orientation is important in watching activities, less so in listening activities; posture is relatively less important in activities which do not require active responses, such as silent reading, and more important in activities which do require periodic responding, such as taking notes or writing.

Momentary lapses of attention (i.e., orienting responses) are not rated as unproductive if the child is engaged in relevant activity and returns to it immediately. The convention is to rate the interval as unproductive if the attention lapse occupies half of the ten second interval or more.

Redundancy Ratings

Some behaviors produce one kind of environmental effect if emitted singly but another kind if emitted at frequent intervals. For example, a child who sharpens his pencil only once during a seatwork period would probably be acting in a relevant manner; four or five trips to the pencil sharpener, however, would be considered unproductive at best. The rating convention is to classify the first such behavior as relevant and subsequent episodes within the same observation period as unproductive--or disruptive, if that indeed is the effect. The probable effect of an operant may vary as a function of duration, too. A common example is hand-raising. The child who requests help by raising his hand is exhibiting appropriate behavior in most classrooms. If the hand stays up for an appreciable time, the behavior becomes increasingly less relevant. If 25 or 30% of the working repertoire consists of signalling for attention, the child is not making the best use of his time. The convention in this case is to rate the first full interval of each hand-raise as relevant and subsequent intervals within the same movement cycle as unproductive.

Sticks and Stones

For rating purposes, two basic kinds of aggressive behavior are recognized: (1) behavior that interferes with or is harmful to other people, and (2) behavior that withholds from other people the control of their property, or results in its damage or destruction. The first category involves physical

contact or an audible threat of imminent punishment. It is recognized that there are other kinds of aggression against the person--verbal aggression, for example--but the problems of assessing its many forms are so great that Function Rater judgments are tied to more concrete phenomena. A measure of support for the validity of this approach is found in the schoolboy admonition,

> "Sticks and stones may break my bones,
> but names will never hurt me."

Sticks and stones are rated Aggressive; names are classified as Disruptive. Audible threats should be rated Aggressive if, in the judgment of the rater, the threat is of sufficient intensity to represent a real hazard to the threatened party--something he must contend with by fighting or backing down. It is not intended that purely verbal behavior--bantering back and forth about who is going to do what to whom--be classified as aggressive or that threats of future retribution ("I'll get you after school") be so judged. Intensity and imminence of harm are the key factors in audible threats.

The second form of Aggressive behavior centers on actions against property. Elaborate thefts will seldom be observed during behavior ratings. More likely a child will be seen taking an object away from another--a pencil from his hand, a hat from his head, an object brought for "show and tell." In any case where the property is forcibly wrested from another, it is rated Aggressive. The rater may subsequently hear that the act was in retaliation for a similar misdeed previously committed, or that the offender was merely trying to recover his own property. No matter, the segment in which it occurs is marked Aggressive. It should be remembered that the purpose of the rating system is to describe behavior in gross quantitative terms, irrespective of its causes or justifications. When such information is available, however,

it can be included in the comments section of the Rater form. Sometimes a child will borrow the property of another, simply helping himself without bothering with the amenities of obtaining permission. This should not be judged Aggressive unless it excites an unrelieved protest from the offended party, or if, after taking unguarded property, a child conceals it or passes it to another for concealment.

Disruptive behavior is easy to recognize but difficult to describe. In the matter of gaining attention, children often show singular energy and creativity; hence, the possibilities are virtually limitless. In general, disruptive behavior is characterized by motion or sound that interrupts the teacher-directed focus of attention through distraction of other children or the teacher herself, or would tend to do so in a typical classroom.

Typical Classroom

It was stated in the main text that ratings are made with respect to the social context in which they occur. The intent is to make allowances for the broad variety of classroom rules that are in effect from one classroom to another and even within the same class at different times. For example, a teacher may allow no talking at all during an arithmetic seatwork period, but approve normal conversation between children who are working together in a science project. The rater can usually deduce the rule structure of a classroom in one or two observation periods. Occasionally, however, a class is observed in which there appears to be no enforced rule for acceptable conduct. If there is general chaos, it is no longer appropriate to judge an individual child's behavior in relation to its social context. In this situation, the rule of Typical Classroom is invoked, and behaviors that would be disruptive in most classrooms are so rated. This convention holds even though the

behavior of those whom the rated child would disrupt is worse than his own. On this point, it should be noted that the rating of an individual child is relatively meaningless unless there is a group rating of his classmates with which to compare. A child with 45% disruptive and 10% aggressive behavior may sound demonic, but if this is the class norm the focus of any intervention would shift from the individual to the structure of the classroom in general.

TABLE I

Summary of relative frequencies of relevant, unproductive, disruptive and aggressive behaviors observed in regular and special class settings

| Location and Time | Label and Grade | Pupils Rated | Days Rated | Raters | Total Intervals Rated | Inter-Rater Agreement Agree | Disagree | % | Relative Frequencies in Percents Rel. | Unpr. | Dis. | Agg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Regular Class Spring, 1969 | Normal Gr. 5 | 61** | 15 | 1 | 2087 | | | - | 63.5 | 26.1 | 10.2 | .3 |
| Special School Summer, 1972 | TMR Primary | 12** | 14 | 2 | 770 | 687 | 18 | 97.4 | 53.5 | 32.8 | 13.1 | .5 |
| Middle Schools Fall, 1972 | Slow Learners Gr. 6, 7 | 27* | 31 | 7 | 4050 | | | - | 60.8 | 27.9 | 11.0 | .3 |
| Token Economy Spring, 1972 | SEM, LD EMR Intermediate | 9* | 20 | 14 | 1932 | 1691 | 166 | 91.1 | 72.2 | 19.7 | 7.5 | .6 |
| Open Classroom Fall, 1972 | Slow Learners Gr. 6, 7 | 45* | 61 | 5 | 4868 | | | - | 71.9 | 23.0 | 4.6 | .5 |
| Departmenta-lized Special Class 1972-73 | SEM, LD EMR Gr. 1-5 | 14** | 25 | 3 | 5575 | 663 | 42 | 94.0 | 70.0 | 24.6 | 5.2 | .2 |
| Regular Open Spring, 1972 | Normal Gr. 1-5 | 5* | 5 | 2 | 500 | 477 | 23 | 95.4 | 66.8 | 16.6 | 13.4 | 3.2 |
| Special Class Spring, 1972 | SEM Intermediate | 4* | 4 | 2 | 250 | 222 | 20 | 91.8 | 65.2 | 22.4 | 11.2 | 1.2 |
| Grand Totals | | 177 | 175 | 36 | 20032 | 4694 | 332 | 93.4 | 67.4 | 24.7 | 7.5 | .4 |

* Individual ratings combined

** Group rating mode

FIGURE 1

DATE 4-17-74

DELAWARE FUNCTION RATER

NAME Reading Group (5 kids x 4 min.)

Row _____
Seat _____

R = Relevant
U = Unproductive
D = Disruptive
A = Aggressive

| | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | Comments (50-60) | R | U | D | A |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | R | U | U | D | R | | 2 | 2 | 1 | |
| 2 | R | R | R | R | R | | 5 | | | |
| 3 | R | R | U | U | U | | 2 | 3 | | |
| 4 | U | U | R | R | R | | 3 | 2 | | |
| 5 | R | R | R | R | R | | 5 | | | |
| 6 | R | R | U | U | U | hand raised | 2 | 3 | | |
| 7 | U | U | U | U | U | gives up | | 5 | | |
| 8 | R | R | R | R | R | | 5 | | | |
| 9 | R | R | R | R | R | | 5 | | | |
| 10 | U | R | R | R | R | | 4 | 1 | | |
| 11 | R | R | U | U | D | | 2 | 2 | 1 | |
| 12 | D | D | R | R | R | | 3 | | 2 | |
| 13 | R | R | R | R | R | | 5 | | | |
| 14 | R | R | R | D | A | swats neighbor | 3 | | 1 | 1 |
| 15 | R | D | D | D | U | | 1 | 1 | 3 | |
| 16 | U | U | U | R | R | | 2 | 3 | | |
| 17 | R | R | R | R | R | | 5 | | | |
| 18 | R | R | R | R | R | | 5 | | | |
| 19 | U | U | U | U | U | | | 5 | | |
| 20 | U | U | D | D | U | | | 3 | 2 | |

Rater Mike G.

Class and Teacher Miss P.

School Anonymous

| | R | U | D | A |
|---|---|---|---|---|
| Total | 59 | 30 | 10 | 1 |
| Percent | | | | |